Characterization of β-casein Gene Sequence Variants in Cholistani Cattle

Ruqayya Bint Khalid¹, Asif Nadeem¹,²*, Maryam Javed¹, Muhammad Zubair Shabbir³ and Masroor Ellahi Babar⁴

¹Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore, Pakistan.
²Department of Biotechnology, Virtual University of Pakistan, Lahore, Pakistan
³Department of Microbiology, University of Veterinary and Animal Sciences, Lahore, Pakistan
⁴The University of Agriculture, Dera Ismail Khan, Khyber Pakhtunkhwa, Pakistan

A B S T R A C T

β-casein is second most abundant protein of cow's milk. β-casein gene is highly polymorphic. A1 and A2 are the frequently occurred variants. A1 is recognized as potential cause of several human diseases. It is important to evaluate the A1/A2 β-casein status in milk. Current study was conducted to molecular characterize the exonic regions of β-casein gene and to explore the status of A1/A2 β-casein type in Cholistani cattle breed of Pakistan. Blood samples of Cholistani Cattle were collected from Government Livestock Farm, Jugait Peer, Bahawalpur. Genomic DNA was extracted from whole blood by organic method. PCR primers were designed and optimized according to respective melting temperatures. PRALINE tool, MEGA 6.0 and POPGENE tool were utilized for phylogenetic tree construction and statistical analysis, respectively. Characterization of physical and chemical properties of β-casein protein was performed by ProtParam and SWISS MODEL was utilized for protein model prediction. Sequencing results of amplified PCR products revealed total 9 SNPs including 1 in exonic and 8 in intronic regions of β-casein gene. Results represent the presence of A2 allele in milk of Cholistani Cattle of Pakistan. Multiple sequence alignment represented the presence of MKVLILACLVALALARE and QRAVPVQALLLNQE as highly conserved regions. Phylogenetic analysis revealed the evolutionary relationship among Cholistani cattle of Pakistan, Bos indicus and Bos taurus. β-casein gene is highly polymorphic and A2 allele is present in Cholistani Cattle of Pakistan.

INTRODUCTION

Agriculture is considered as an important sector of Pakistan's economy. Livestock species play a vital role in country's economic, social and cultural values. It contributes about 58.3% to the agriculture sector and 11.4% to the overall GDP of Pakistan. Within livestock sector, milk is the major commodity. Annual milk production of Pakistan is over 50 billion liters (GOP, 2018). Milk is considered as a complete diet as it is composed of essential micro and macronutrients. Milk is an essential source of nutrients like proteins, carbohydrates and particular micronutrients. It is rich in vitamin B12, vitamin A, riboflavin, folate, and calcium but iron occurs in less amount (Bermudez *et al*., 2010).

Proteins are one of the most important component of milk, out of which caseins have gained most of the attention due to their recognized association with health related properties. Bovine milk contains four casein components including αS1 (39-46%), αS2 (8-11%), β-casein (25-35%), κ-casein (8-15%) respectively (Roginski *et al*., 2003). β-casein is the 2ⁿᵈ most abundant protein with all essential 209 amino acids. β-casein is highly polymorphic milk protein gene with 12 known variants including A1, A2, A3, B, C, D, E, F, H1, H2, I and G. However, A1 and A2 are the most common occurring variants (Farrell Jr *et al*., 2004; Kamiński *et al*., 2007).

Bovine β-casein gene is located at chromosome number 6, with total length of 8,695 bp and 8 number of exons (NC_037333.1). A1 and A2 allelic variance occur at exon VII and exon VII is responsible for encoding most of the mature protein parts (Ganguly *et al*., 2013). A1 differs from A2 variant in such a way that A1 milk contain histidine (CAT) amino acid at position 67 while A2 milk has proline (CCT) at similar position. This polymorphism leads to conformational changes in the secondary structure of protein and affect casein micelles physical properties. After substitution of histidine, A1 variant becomes vulnerable to gastrointestinal proteolysis digestion causing the release of a 7 amino acid bioactive

peptide, betacasomorphin-7 (BCM-7) (Cieślińska *et al.*, 2007; De Noni, 2008). Several studies have reported adverse effect of BCM-7 on human health like diabetes type I (Elliott *et al.*, 1999; Laugesen and Elliott, 2003), heart diseases (Laugesen and Elliott, 2003; McLachlan, 2001), sudden death syndrome in infants (SIDS) (Sun *et al.*, 2003), atherosclerosis (Tailford *et al.*, 2003) and autism (Cieślińska *et al.*, 2015). Due to association with several human diseases, A1 has been titled as "Devil in the milk". Prevalence of A1 and A2 variants differ among cattle species. A1 type is common in European cattle milk while A2 variant is secreted in milk of Asian and African cattle (Woodford, 2009). Sequence of animals other than cattle and human β-casein gene shows the presence of A2 variant (Ng-Kwai-Hang and Grosclaude, 2003).

Due to adverse effect of A1 variant on human health, screening of β-casein gene status in cattle has gained importance from breeding point of view. While no significant evidence has been reported regarding association of A2 allele with certain diseases, therefore, A2A2 genotype gained importance in breeding programs. Polymorphism in β-casein gene is highly associated with milk performance traits. These variations might help in selection of cattle with better milk quality and yield. Keeping in view the importance of β-casein gene, current study was conducted to molecular characterize the exonic regions of β-casein gene and to explore the status of A1/A2 β-casein type in Cholistani cattle breed of Pakistan.

MATERIALS AND METHODS

Sampling

Blood samples of Cholistani cattle breed (n=20), from Government Livestock Farm, Jugait Peer, Bahawalpur, were collected in EDTA vacutainers. Samples were transported in laboratory in ice box. Samples were stored at 4°C to avoid any degradation.

PCR amplification of β-casein

Genomic DNA was extracted from whole blood using standard phenol chloroform extraction protocol (Maniatis, 1989).

Specific primers were designed against exonic regions β-casein gene for PCR, by using Primer3 (version 0.4.0) software. Reference sequence of *Bos taurus* (NC_037333.1) reported in NCBI database was used for primer designing. For further confirmation and optimization PCR Primer stats (Stothard, 2000), OligoCalc (Kibbe, 2007) were also used. UCSC genome browser (Kent *et al.*, 2002) was used to perform In Silico PCR to check the specificity of primers.

PCR reaction mixture of 25 µl volume was prepared by adding 0.5 µl of each primer (10pmol), 0.25 µl genomic DNA (50ng/ µl), 2.5 µl PCR buffer (10X), 2.5 µl dNTPs (2.5mM/µl), 2 µl MgCl$_2$ (2.5M/µl), 0.3 µl *Taq* DNA Polymerase (5U/µl) and 16.45 µl deionized water. PCR amplification was achieved under these conditions: initial denaturation at 95°C for 5 min; followed by 10 cycles of 94°C for 30 s, 64°C for 30 s and 72°C for 30 s; thereafter 30 cycles of 94°C for 30 s, 54°C for 30 s and 72°C for 30 s and a final extension 72°C for 10 min. Amplified PCR products were evaluated in 1.5% agarose gel. The purified PCR samples were sequenced by using BigDye Terminator on a Genetic Analyzer (ABI 3130XL). These products were sequenced by using the method of Sanger sequencing. Sequenced data was analyzed with the help of BioEdit software (Hall, 1999).

Sequence retrieval and multiple sequence alignment

Nucleotide and protein sequences of β-casein of eight different species were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/). These species include *Bos taurus* (NC_037333.1), *Bos indicus* (NC_032655.1), *Mus musculus* (NC_000071.6), *Capra hircus* (NC_030813.1), *Ovis aries* (NC_040257.1), *Equus caballus* (NC_009146.3), *Oryctolagus cuniculus* (NC_013683.1) and *Homo sapiens* (NC_000004.12), respectively.

To identify the conserved regions of β-casein protein among all these species, multiple sequence alignment of protein sequences was performed with the help of PRALINE tool (Simossis and Heringa, 2005) by using BLOSUM62 exchange weights matrix, associated gap open penalty of 12 and associated gap extension penalty of 1, respectively. PSI-BLAST iterations were set at 3 at an e-value cut off of 0.01.

Phylogenetic tree construction

NJ (Neighbor-Joining) tree was constructed by using MEGA 6.0 software (Tamura *et al.*, 2013). Model selected for NJ tree construction was maximum composite likelihood model and pair wise deletion was selected for gap/missing data treatment. Phylogenetic relationships among the nucleotide sequences of β-casein of species under investigation were based on genetic distances. Total number of 1000 bootstrap replications were selected to calculate the reliability of tree by bootstrap confidence values, respectively. All of the nucleotide sequences were trimmed to equal length, so that, they all correspond to same regions before the generation of tree.

Estimation of evolutionary divergence between sequences

Analysis for the estimation of evolutionary divergence between sequences were conducted by selecting the maximum composite likelihood model and 1000 bootstrap iterations were selected to get the standard error estimates

(Tamura *et al*., 2004). This analysis involved nine nucleotide sequences. This analysis was performed with the help of MEGA6 software (Tamura *et al*., 2013).

Protein structure prediction

SWISS-MODEL was used for the prediction of 3D structure of β-casein protein. Both wild and mutated protein structures were predicted in order to identify the change that occurred in structure of protein due to amino acid change (Waterhouse *et al*., 2018).

Physical and chemical properties identification

Physical and chemical parameters of β-casein protein of Cholistani cattle were identified by using ProtParam tool available on ExPASy proteomics tools (Gasteiger *et al*., 2005).

Statistical data analysis

POPGENE tool (Yeh *et al*., 1997) was used for calculation and identification of allelic and genotypic frequencies as well as for the calculation of Hardy-Weinberg equilibrium. Allelic and genotypic frequencies significance verification was done with the help of Chi-square test.

RESULTS

Purpose of this study was to molecular characterize the exonic regions of β-casein gene and to explore the status of A1/A2 β-casein type in Cholistani cattle breed of Pakistan. BioEdit software was used for identification of variations in the sequence. Total 9 SNPs were identified including 1 in exonic and 8 in intronic regions of β-casein gene (Table I). The exonic SNP was non-synonymous as arginine changed to serine (Table II). In this study, A1/A2 allelic polymorphism in cholistani cattle was also being explored. Our results represent the presence of β-casein A2 allele in milk of Cholistani Cattle of Pakistan (Fig. 1).

Table I. List of identified polymorphisms in β-Casein gene.

Sr. No	SNP ID	Chromosomal position	Nucleotide change	Intronic/ Exonic
1	*CSN* 1	85455593	T > T/C	Intronic
2	*CSN* 2	85455721	T > T/C	Intronic
3	*CSN* 3	85451132	C > G	Exonic
4	*CSN* 4	85450071	C > C/T	Intronic
5	*CSN* 5	85450356	T > T/G	Intronic
6	*CSN* 6	85450012	A > A/C	Intronic
7	*CSN* 7	85450410	C > A	Intronic
8	*CSN* 8	85450362	A > A/G	Intronic
9	*CSN* 9	85450382	T > T/G	Intronic

Table II. Non-synonymous substitutions of amino acids in exonic regions of β-casein gene.

SNP ID	Chromosomal position	Nucleotide change	Amino acid change	Consequence type	Amino acid coordinates
CSN 3	85451132	C > G	R>S	Missense variant	129

Multiple sequence alignment of nine protein sequences was done using PRALINE tool. MKVLILACLVALALARE and QRAVPVQALLLNQE were found to be highly conserved regions among various species (Fig. 2). Phylogenetic tree was constructed with the help of MEGA6 software to find out the evolutionary relationships among species. Figure 3 represents the optimal tree with the sum of branch length 1.90855609. The percentage of replicate trees are shown next to the branches (associated taxa are clustered together in bootstrap test). NJ tree was drawn to scale, having similar units of branch lengths as those of evolutionary distances utilized to deduce the phylogenetic tree. This tree represents that Cholistani cattle is closely related to *Bos indicus* and *Bos taurus*.

Pairwise distances were measured by using MEGA6 software. Total number of base substitutions per site from between sequences and standard error estimate(s) above the diagonal have been shown in Table III. Here the lower diagonal represents the average genetic distances between species and upper diagonal shows the standard error estimate(s). Evolutionary divergence estimation indicates that the Cholistani cattle is highly similar to *Bos taurus* and *Bos indicus* and it is highly divergent from *Mus musculus* and *Homo sapiens*, respectively. As non-synonymous mutations lead to change in amino acid residue. SWISS MODEL was utilized for indication of change in structure of β-casein protein due to amino acid change (Fig. 4).

Physical and chemical parameters of β-casein protein were calculated with the help of ProtParam tool. Table IV represents the physiochemical properties of β-casein protein of Cholistani cattle. Isoelectric point represents the acidic nature of protein. GRAVY (Grand average of hydropathicity) and aliphatic index are usually correlated. GRAVY is used to predict water and protein interaction. Lower GRAVY value of Cholistani β-casein protein indicates the hydrophilic nature of protein. Allelic frequencies were calculated with the help of POPGENE. Highest allelic frequency of A2 allele was 0.8125 while lowest calculated allelic frequency was 0.3750 (Table V). Wild and mutant genotypic frequencies were calculated and has been shown in Table VI. POPGENE tool was utilized for the calculation of genic variation (Table VII) and heterozygosity of all loci (Table VIII).

Table III. Evolutionary divergence between species.

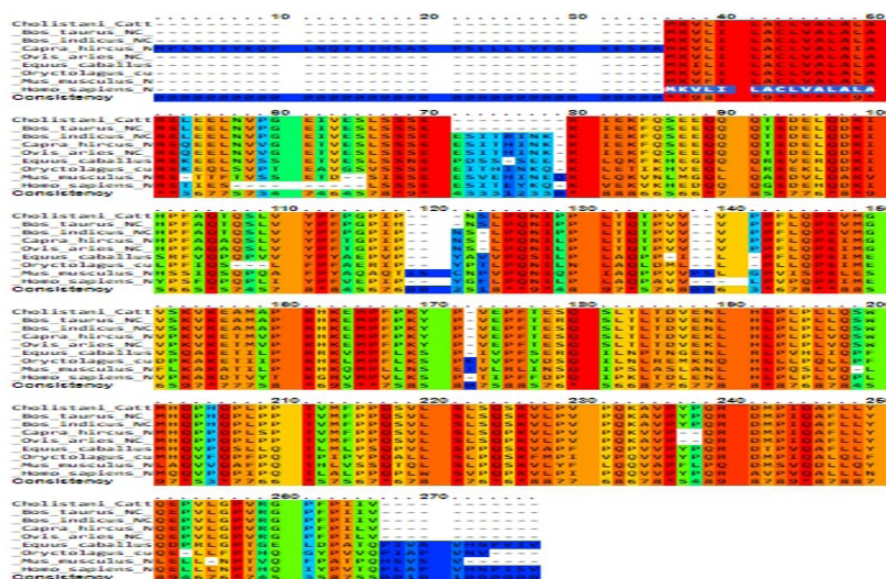|	1	2	3	4	5	6	7	8	9
Cholistani Cattle_Pakistan		0.000	0.027	39.697	9.180	0.279	17.881	25.182	35.275
Bos taurus NC_037333.1	0.000		0.027	39.696	2.057	1.179	17.881	25.182	35.275
Bos indicus NC_032655.1	0.056	0.056		40.610	130.391	131.702	17.871	25.303	36.094
Mus musculus NC_000071	20.231	20.231	20.701		41.909	41.815	42.208	44.042	40.813
Capra hircus NC_030813.1	0.214	0.215	0.248	21.285		0.006	19.030	26.422	35.840
Ovis aries NC_040257.1	0.217	0.218	0.250	21.263	0.016		19.072	26.453	35.934
Equus cabalus NC_009146.3	8.472	8.741	8.720	21.514	9.364	9.410		35.748	39.043
Oryctolagus curiculus NC__013683.1	12.674	12.674	12.744	22.083	13.326	13.342	17.984		36.724
Homo sapiens NC_000004.12	17.836	17.836	18.287	20.774	18.229	18.290	19.782	18.589	

Fig. 1. At position 85451290 of β-casein gene, AGG codon presence confirms the presence of A2 allele in cholistani cattle of Pakistan.

Fig. 2. Multiple sequence alignment of nine β-casein protein sequences of different species.
Key: Conserved 012345678910 unconseved.

Table IV. Physical and chemical parameters of β-casein protein of Cholistani cattle.

Properties	
Molecular weight (Da)	24165.24
Amino acids	216
Isoelectric point (pI)	5.11
Aliphatic index (AI)	97.36
Instability index (II)	93.74
GRAVY (Grand average of hydropathicity)	-0.123
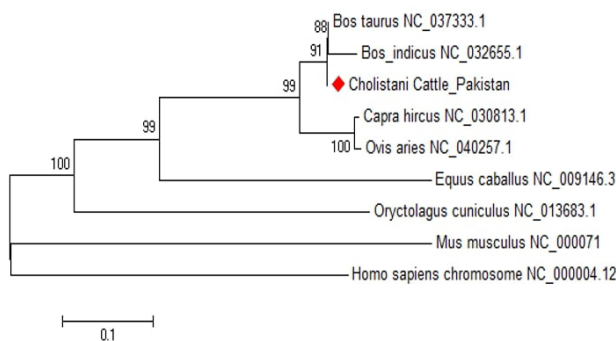
Fig. 3. Phylogenetic tree of β-casein gene of nine species constructed with the help of Neighbor-Joining method with bootstrap of 1000 iterations and Maximum composite likelihood method.
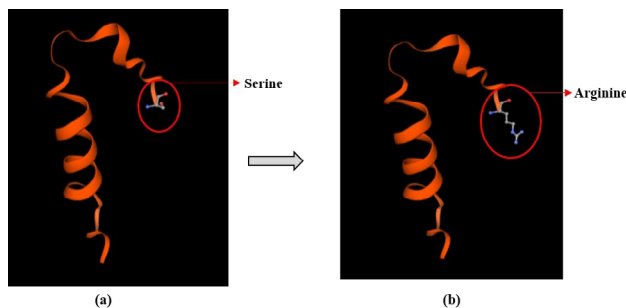
Fig. 4. Protein structure of wild β-casein protein (a), mutant β-casein protein (b).

DISCUSSION

Milk is a complete food as it contains all the basic nutrients for consumption. However, milk proteins remain the center of controversy in regard their association with human health issues. As several studies have reported the association of BCM-7, bioactive peptide released after proteolytic digestion of A1 milk, with several diseases (Laugesen and Elliott, 2003; McLachlan, 2001; Sun et al., 2003; Tailford et al., 2003). β-casein gene is highly

polymorphic and A1/A2 allelic variance of β-casein gene is considered as most important polymorphic variants. This study was conducted to molecular characterize the exonic regions of β-casein gene and to explore the status of A1/A2 β-casein type in Cholistani cattle breed of Pakistan.

Table V. Allelic frequencies of polymorphisms identified.

Sr. No.	SNP ID	SNP position	Allelic Frequency		X^2 test (p-value)
			A	B	
1	CSN 1	85455593	0.6042	0.3958	0.000191
2	CSN 2	85455721	0.6667	0.3333	0.001450
3	CSN 3	85451298	0.3333	0.6667	0.000040
4	CSN 4	85450071	0.6042	0.3958	0.000191
5	CSN 5	85450356	0.6875	0.3125	0.090921
6	CSN 6	85450012	0.3750	0.6250	0.000034
7	CSN 7	85450410	0.6458	0.3542	0.000233
8	CSN 8	85450362	0.6875	0.3125	0.000312
9	CSN 9	85450382	0.6250	0.3750	0.001069

Table VI. Genotypic frequencies of identified polymorphisms.

Sr. No.	SNP ID	SNP position	Genotypic frequency			X^2 test (p-value)
			AA	AB	BB	
1	CSN 1	85455593	0.54	0.12	0.33	0.000191
2	CSN 2	85455721	0.54	0.16	0.29	0.001450
3	CSN 3	85451298	0.33	0.08	0.58	0.000040
4	CSN 4	85450071	0.54	0.12	0.33	0.000191
5	CSN 5	85450356	0.50	0.29	0.16	0.090921
6	CSN 6	85450012	0.33	0.08	0.54	0.000034
7	CSN 7	85450410	0.54	0.12	0.33	0.000233
8	CSN 8	85450362	0.58	0.08	0.29	0.000312
9	CSN 9	85450382	0.54	0.20	0.25	0.001069

A total of nine single nucleotide polymorphisms have been identified in β-casein gene of Cholistani cattle breed of Pakistan. Eight identified SNPs were found in intronic region while one was present in exonic region. C allele serves as a genetic marker for A2 variant of β-casein (Caroli et al., 2009). It has been reported that SNP at 85451132 (CSN 3) position, β-casein C allele helps in discrimination of B and C allele from A1 variant (Cecchinato et al., 2014). It was reported that β-casein 85451132 C allele was positively associated with curd firming (Cecchinato et al., 2015). C > G polymorphism has also been found

at 85451132 (*CSN* 3) chromosomal position of Cholistani cattle. 85450410 (C>A) have also been reported in Holstein crossbred while studying association between protein composition traits and polymorphisms in milk protein genes and SNPs were significantly associated with milk protein composition traits (Huang *et al.*, 2012).

Table VII. Genic variation statistics for all loci.

Sr. No.	SNP ID	*na	*ne	*I
1	*CSN* 1	2.00	1.9168	0.6713
2	*CSN* 2	2.00	1.8000	0.6365
3	*CSN* 3	2.00	1.8000	0.6365
4	*CSN* 4	2.00	1.9168	0.6713
5	*CSN* 5	2.00	1.7534	0.6211
6	*CSN* 6	2.00	1.8824	0.6616
7	*CSN* 7	2.00	1.8432	0.6500
8	*CSN* 8	2.00	1.7534	0.6211
9	*CSN* 9	2.00	1.8824	0.6616
	Mean	2.00	1.8063	0.6341
	St. Dev	0.00	0.1358	0.0535

* na, Observed number of alleles; * ne, Effective number of alleles [Kimura and Crow (1964)]; * I, Shannon's Information index [Lewontin (1972)].

Table VIII. Heterozygosity statistics for all loci.

Sr. No.	SNP ID	Obs_Hom	Obs_Het	Exp_Hom*	Exp_Het*	Nei**	Ave_Het
1	*CSN* 1	0.8750	0.1250	0.5115	0.4885	0.4783	0.4783
2	*CSN* 2	0.8333	0.1667	0.5461	0.4539	0.4444	0.4444
3	*CSN* 3	0.9167	0.0833	0.5461	0.4539	0.4444	0.4444
4	*CSN* 4	0.8750	0.1250	0.5115	0.4885	0.4783	0.4783
5	*CSN* 5	0.7083	0.2917	0.5612	0.4388	0.4297	0.4297
6	*CSN* 6	0.9167	0.0833	0.5213	0.4787	0.4688	0.4688
7	*CSN* 7	0.8750	0.1250	0.5328	0.4672	0.4575	0.4575
8	*CSN* 8	0.8750	0.1250	0.1250	0.1250	0.4297	0.4297
9	*CSN* 9	0.8333	0.1667	0.5213	0.4787	0.4688	0.4688
Mean		0.8409	0.1591	0.5475	0.4525	0.4430	0.4430
St. Dev		0.0928	0.0928	0.5475	0.0502	0.0492	0.0492

* Expected homozygosty and heterozygosity were computed using Levene (1949); ** Nei (1973) expected heterozygosity.

Single amino acid substitution in A2 allele of β-casein gene is one of the most common and abundant form of genetic polymorphism that results in the production of A1 allele. These variations can be helpful for germplasm fingerprinting and population studies. Insilico analysis of β-casein gene indicated that A1 and A2 allele differ only at 67th position, where histidine amino acid is substituted instead of proline (Jaiswal and Sarsavan, 2013).

Presence of A2 allele in Cholistani cattle of Pakistan can be reported in this study. Similarly, it was reported that local Ladakh cattle in India have higher A2 allelic frequency and exhibit A2A2 genotype which is the source A2 milk (Sharma *et al.*, 2018). A relatively high allelic frequency of A1 β-casein variant was identified in Iranian naive cattle breed (Gholami *et al.*, 2016). Another study on Bangladesh bovine reported high A2A2 genotype, whereas, no A1A1 genotype was found. Heterozygous A1A2 genotype was also present in some breeds. So, it was concluded that A2 β-casein variants were dominant in Bangladesh.

Major milk proteins, including casein, have been found to perform important biological activities. These proteins are involved in the regulation of various health conditions of human beings as well as involved in provision of immunological protection and important nutritive components. Presence of A2 allele was reported in Sahiwal cattle breed, responsible for the prevention of harmful BCM-7 opioid production after enzymatic digestion of milk in stomach. It was also reported that crossbred cattle breed composed of A1 genetic variant while indigenous cattle produce A2 β-casein variant milk (Kumar *et al.*, 2019). A2 β-casein gene variant has also been found in this study.

Various studies have reported that consumption A1 β-casein gene variant milk was correlated with high mortality rate. A2 β-casein variant is good for health as compared to A1 variant (Laugesen and Elliott, 2003). In a study conducted on sahiwal cattle breed of Pakistan population, A2 allelic variance was reported, similar to results of this study (Mir *et al.*, 2014). Whereas, A1A1 and A1A2 genotypes of β-casein gene were detected in Holstein and native cattle breeds (Firouzamandi *et al.*, 2018).

Association of genetic variants of casein and milk serum proteins with milk, protein, and fat production was studied in dairy cattle. Strong association of B allele of αsl- casein gene and A allele of β-casein gene with high protein, fat and milk yields was reported. β-casein variants were found to be correlated with fat percentage as well as fat yield. It was also reported that A3 allele was highly associated with milk production trait (Ng-Kwai-Hang *et al.*, 1984). Polymorphisms in β-casein gene are highly associated with performance traits. It was observed that β-casein gene (as a member of casein cluster) influence the traits regarding milk performance (Singh *et al.*, 2015). A study was designed to find out the effects of genetic polymorphisms of caseins on Swedish

Red, Danish Holstein, and Danish Jersey cattle breeds. Milk composition differ in each breed significantly. These variations might help in selection of cattle with better milk quality and yield. These might also help in processing of cheese (Gustavsson *et al*., 2014).

Multiple sequence alignment of protein sequences represented highly conserved region MKVLILACLVALALARE. Similar region has been found conserved in another study (Pareek *et al*., 2012). Phylogenetic analysis revealed the evolutionary relationship of β-casein among Cholistani Cattle_Pakistan, *Bos indicus* and *Bos taurus*. Similarly, cattle, goat and sheep formed a clade, as they all belong to same *bovidae* family. Rabbit and horse gene sequences were somewhat related to this clade but maintained at a distant relationship. These findings are similar with some other studies (Pareek *et al*., 2012; Vincent *et al*., 2014).

Physical and chemical analysis of Cholistani cattle protein indicated that this protein is acidic in nature. Aliphatic index represented the stability of proteins at high temperature. Grand average hydropathicity (GRAVY) value of this protein indicated the hydrophilic nature of proteins. This might be helpful to increase the oligomerization as well as functional properties of proteins. Similarly, in another study aliphatic index, GRAVY, and hydrophobicity all suggested αs1-CN as a moderately hydrophobic protein (Huppertz *et al*., 2018).

Genotypic frequencies calculated in Cholistani Cattle for homozygous A2A2 waere 0.62 and for heterozygous A1A2 was 0.37, respectively. Genetic polymorphisms of β-casein gene were studied in various breeds of cattle. β-casein have been found to constitute approximately 45 percent of bovine milk caseins. A1 and A2 are the most common whereas, B is the less common allele of β-casein gene in dairy cattle. A1 was found to be associated with autism, type I diabetes mellitus and coronary heart diseases whereas, A2 variant was found to be associated with reduction of serum cholesterol. Genotypic frequencies detected for homozygous A1A1 genotype was 0.1261, for A2A2 homozygous 0.5405. For heterozygous A1A2 genotype frequency in Simmental breed frequency was found to be 0.3333. Similarly, 0.1379, 0.4023 and 0.4598 genotypic frequencies for A1A1, A2A2 and A1A2 were detected for Holstein breed and 0.3034, 0.1798 and 0.5168 for Pinzgau breed, respectively. A2 allelic frequency was higher in Holstein and Simmental breeds respectively whereas, with 0.5618 A1 allelic frequency was higher in Pinzgau breed (Miluchová *et al*., 2014). A2A2 and A2B genotypic frequencies of Sahiwal cattle were 0.14 and 0.86, respectively (Mir *et al*., 2014).

Heterozygosity statistics for all loci of Cholistani cattle were calculated using Levene's heterozygosity and homozygosity and Nei's (1973) expected heterozygosity. Mean value for Nei's calculation was 0.4430. In another study, genetic variation of β-casein gene was find out. Heterozygosity statistics revealed that Holstein population had 0.4983 Nei's (1973) expected heterozygosity. Sarabi population had 0.4990 and Gaja population had 0.5000 Nei's expected heterozygosity value (Firouzamandi *et al*. 2018). Genic variation statistics for all loci of Cholistani cattle was calculated using Shannon's information index. Similar method was also utilized in another study finding genetic diversity of β-casein gene in Indian goats (Rout *et al*., 2010).

CONCLUSION

Hence, sequence characterization has shown that β-casein gene is highly polymorphic with the identification of exonic and intronic single nucleotide polymorphisms. Present study results indicated that A2 β-casein variant is present in Cholistani cattle breed of Pakistan. Phylogenetic analysis revealed the species relatedness that is consistent with well-established evolutionary history. Dairy cattle breeders should start genotyping of all cattle for A2A2 β-casein variants. Selective breeding programs should be improved to gain good quality and quantity milk.

Statement of conflict of interest
The authors have declared no conflict of interest.

REFERENCES

Bermudez, O.I., Toher, C., Montenegro-Bethancourt, G., Vossenaar, M., Mathias, P., Doak, C. and Solomons, N.W., 2010. Research dietary intakes and food sources of fat and fatty acids in guatemalan school children: A cross-sectional study. *Nutri. J.,* **9**: Article number: 20. https://doi.org/10.1186/1475-2891-9-20

Caroli, A., Chessa, S. and Erhardt, G., 2009. Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* **92**: 5335-5352. https://doi.org/10.3168/jds.2009-2461

Cecchinato, A., Chessa, S., Ribeca, C., Cipolat-Gotet, C., Bobbo, T., Casellas, J. and Bittante, G., 2015. Genetic variation and effects of candidate-gene polymorphisms on coagulation properties, curd firmness modeling and acidity in milk from brown Swiss cows. *Animal*, **9**: 1104-1112. https://doi.org/10.1017/S1751731115000440

Cecchinato, A., Ribeca, C., Chessa, S., Cipolat-Gotet, C., Maretto, F., Casellas, J. and Bittante, G., 2014. Candidate gene association analysis for milk yield,

composition, urea nitrogen and somatic cell scores in brown Swiss cows. *Animal*, **8**: 1062-1070. https://doi.org/10.1017/S1751731114001098

Cieślińska, A., Kamiński, S., Kostyra, E. and Sienkiewicz-Szłapka, E., 2007. Beta-casomorphin 7 in raw and hydrolyzed milk derived from cows of alternative β-casein genotypes. *Milchwissenschaft*, **62**: 125-127.

Cieślińska, A., Sienkiewicz-Szłapka, E., Wasilewska, J., Fiedorowicz, E., Chwała, B., Moszyńska-Dumara, M., Cieśliński, T., Bukało, M. and Kostyra, E., 2015. Influence of candidate polymorphisms on the dipeptidyl peptidase IV and μ-opioid receptor genes expression in aspect of the β-casomorphin-7 modulation functions in autism. *Peptides,* **65**: 6-11. https://doi.org/10.1016/j.peptides.2014.11.012

De Noni, I., 2008. Release of β-casomorphins 5 and 7 during simulated gastro-intestinal digestion of bovine β-casein variants and milk-based infant formulas. *Fd. Chem.*, **110**: 897-903. https://doi.org/10.1016/j.foodchem.2008.02.077

Elliott, R., Harris, D., Hill, J., Bibby, N. and Wasmuth, H., 1999. Type I insulin-dependent. diabetes mellitus and cow milk: Casein variant consumption. *Diabetologia*, **42**: 292-296. https://doi.org/10.1007/s001250051153

Farrell Jr, H., Jimenez-Flores, R., Bleck, G., Brown, E., Butler, J., Creamer, L., Hicks, C., Hollar, C., Ng-Kwai-Hang, K. and Swaisgood, H., 2004. Nomenclature of the proteins of cows' milk, Sixth revision. *J. Dairy Sci.,* **87**: 1641-1674. https://doi.org/10.3168/jds.S0022-0302(04)73319-6

Firouzamandi, M., Gholami, M., Toloui, M. and Eshghi, D., 2018. Genetic variation of β-casein gene using AS-PCR and ARMS-PCR techniques in bovine populations. *Russia. J. Genet.,* **54**: 1352-1357. https://doi.org/10.1134/S1022795418110054

Ganguly, I., Kumar, S., Gaur, G., Singh, U., Kumar, A., Kumar, S., Mann, S. and Sharma, A., 2013. Status of β-casein., CSN2. polymorphism in frieswal., HF X Sahiwal crossbred. cattle. *Int. J. Biotechnol. Bioeng. Res*., **4**: 249-256. https://doi.org/10.1016/j.rvsc.2012.09.001

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A., 2005. *Protein identification and analysis tools on the ExPASy server*. Springer. pp. 571-607. https://doi.org/10.1385/1-59259-890-0:571

GOP, 2018. *Pakistan economic survey 2017–18*. Finance Division, Economic Advisor's Wing, Islamabad, Pakistan.

Gholami, M., Hafezian, S., Rahimi, G., Farhadi, A., Rahimi, Z., Kahrizi, D., Kiani, S., Karim, H.,

Vaziri, S., Muhammadi, S. and Veisi, F., 2016. Allele specific-PCR and melting curve analysis showed relatively high frequency of β-casein gene A1 allele in Iranian Holstein, Simmental and native cows. *Cell. Mol. Biol.,* **62**: 138-143. http://doi.org/10.0.57.123/cmb/2016.62.12.23

Gustavsson, F., Buitenhuis, A.J., Johansson, M., Bertelsen, H.P., Glantz, M., Poulsen, N.A., Månsson, H.L., Stålhammar, H., Larsen, L.B. and Bendixen, C., 2014. Effects of breed and casein genetic variants on protein profile in milk from Swedish Red, Danish Holstein, and Danish Jersey cows. *J. Dairy Sci.,* **97**: 3866-3877. https://doi.org/10.3168/jds.2013-7312

Hall, T.A.,1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series.*, **41**: 95-98.

Huang, W., Peñagaricano, F., Ahmad, K., Lucey, J., Weigel, K. and Khatib, H., 2012. Association between milk protein gene variants and protein composition traits in dairy cattle. *J. Dairy Sci.* **95**: 440-449. https://doi.org/10.3168/jds.2011-4757

Huppertz, T., Fox, P. and Kelly, A., 2018. *The caseins: Structure, stability, and functionality.* Elsevier, pp. 49-92. https://doi.org/10.1016/B978-0-08-100722-8.00004-8

Jaiswal, K.P. and Sarsavan, A., 2013. Insilico analysis of single nucleotide polymorphism's., SNP's. in Karan Fries cattle of A1, A2 beta casein gene. *Int. J. Environ. Anim. Conserv*., **2**: 29-34.

Kamiński, S., Cieślińska, A. and Kostyra, E., 2007. Polymorphism of bovine beta-casein and its potential effect on human health. *J. appl. Genet*., **48**: 189-198. https://doi.org/10.1007/BF03195213

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D., 2002. The human genome browser at UCSC. *Genome Res*., **12**: 996-1006. https://doi.org/10.1101/gr.229102

Kibbe, W.A., 2007. OligoCalc: An online oligonucleotide properties calculator. *Nucl. Acids Res*., **35**: 43-46. https://doi.org/10.1093/nar/gkm234

Kimura, M. and Crow, J.F., 1964. The number of alleles that can be maintained in a finite population. *Genetics,* **49**: 725-738.

Kumar, S., Singh, R.V. and Chauhan, A., 2019. Molecular characterization of A1/A2 Beta-casein Alleles in Vrindavani crossbred and Sahiwal cattle. *Indian J. Anim. Res*., **53**: 151-155. https://doi.org/10.18805/ijar.B-3493

Laugesen, M. and Elliott, R., 2003. *Ischaemic heart disease, type 1 diabetes, and cow milk A1* β-casein.

Maniatis, T., 1989. *Molecular cloning. Decontamination of dilute solutions of ethidium bromide*. Cold Spring Harbor Laboratory Press, NY.

Levene, H., 1949. On a matching problem arising in genetics. *Annls Math. Statist.,* **20**: 91–94.

Lewontin, R.C., 1972. The apportionment of human diversity. *Evol. Biol.,* **6**: 381–398

McLachlan, C., 2001. β-casein A1, ischaemic heart disease mortality, and other illnesses. *Med. Hypotheses*, **56**: 262-272. https://doi.org/10.1054/mehy.2000.1265

Miluchová, M., Gábor, M., and Trakovická, A., 2014. Analysis of beta-casein gene.,CSN2. polymorphism in different breeds of cattle. *Sci. Pap. Anim. Sci. Biotechnol.*. **47**: 56-59.

Mir, S.N., Ullah, O. and Sheikh, R., 2014. Genetic polymorphism of milk protein variants and their association studies with milk yield in Sahiwal cattle. *Afr. J. Biotechnol*. **13**: 555-565. https://doi.org/10.5897/AJB2013.13216

Ng-Kwai-Hang, K. and Grosclaud, F., 2003. *Genetic polymorphism of milk proteins*. *Springer*. pp. 739-816. https://doi.org/10.1007/978-1-4419-8602-3_22

Ng-Kwai-Hang, K., Hayes, J., Moxley, J. and Monardes, H., 1984. Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. *J. Dairy Sci.,* **67**: 835-840. https://doi.org/10.3168/jds.S0022-0302(84)81374-0

Nei, M., 1973. Analysis of gene diversity in subdivided populations. *Proc. natl. Acad. Sci. U.S.A.,* **70**: 3321–3323.

Pareek, K., Shah, S., and Arora, A., 2012. Casein protein: phylogeny study of important domestic animals using computational method. *Int. J. Adv. Res. Biosci.*, **2**: 326-337.

Roginski, H., Fuquay, J.W. and Fox, P.F., 2003. *Encyclopedia of dairy sciences*. Volumes 1-4. Academic Press, London.

Rout, P., Kumar, A., Mandal, A., Laloe, D., Singh, S. and Roy, R., 2010. Characterization of casein gene complex and genetic diversity analysis in Indian goats. *Anim. Biotechnol*., **21**: 122-134. https://doi.org/10.1080/10495390903534622

Sharma, A., Bharti, V.K., Kumar, B., Iqbal, M., Rabgais, S., Kumar, P., Giri, A., Kalia, S., Gagoi, D., Sarangi, P.P. and Mukesh, M., 2018. Sequence characterisation and genotyping of allelic variants of beta casein gene establishes native cattle of Ladakh to be a natural resource for A2 Milk. *Defen. Life Sci. J.,* **3**: 177-181. https://doi.org/10.14429/dlsj.3.12574

Simossis, VA. and Heringa, J., 2005. PRALINE: A multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucl. Acids Res*., **33**: 289-294. https://doi.org/10.1093/nar/gki390

Singh, L.V., Jayakumar, S., Sharma, A., Gupta, S.K., Dixit, S., Gupta, N. and Gupta, S., 2015. Comparative screening of single nucleotide polymorphisms in β-casein and κ-casein gene in different livestock breeds of India. *Meta Gene*, **4**: 85-91. https://doi.org/10.1016/j.mgene.2015.03.005

Stothard, P., 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques,* **28**: 1102-1104. https://doi.org/10.2144/00286ir01

Sun, Z., Zhang, Z., Wang, X., Cade, R., Elmir, Z. and Fregly, M., 2003. Relation of β-casomorphin to apnea in sudden infant death syndrome. *Peptides,* **24**: 937-943. https://doi.org/10.1016/S0196-9781(03)00156-6

Tailford, K.A., Berry, C.L., Thomas, A.C. and Campbell, J.H., 2003. A casein variant in cow's milk is atherogenic. *Atherosclerosis*, **170**: 13-19. https://doi.org/10.1016/S0021-9150(03)00131-X

Tamura, K., Nei, M. and Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *PNAS,* **101**: 11030-11035. https://doi.org/10.1073/pnas.0404206101

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol*., **30**: 2725-2729. https://doi.org/10.1093/molbev/mst197

Vincent, S., Momoh. O. and Yakubu. A., 2014. Bioinformatics analysis of beta-casein gene in some selected mammalian species. *Res. Opin. Anim. Vet. Sci*. **4**: 564-570.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C. and Bordoli, L., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucl. Acids Res*., **46**: 296-303. https://doi.org/10.1093/nar/gky427

Woodford, K., 2009. *Devil in the milk: Illness, health and the politics of A1 and A2 milk*. Chelsea Green Publishing.

Yeh, F.C., Yang, R., Boyle, T.B.J., Ye, Z.H., and Mao. J.X., 1997. The userfriendly shareware for population genetics analysis. *Edmon. Biol. Biotechnol. Centre, Univ. Alberta, Canada*, **10**: 295-301.